

A Statistic that Identifies Errant Standard Preparation and Instrument Nonlinearity Demonstrated with Mercury Standards Prepared by Blending NIST Fossil Fuel CRMs of Similar Matrices

Bruce S. MacDonald,^{*,†} John L. Molloy,[‡] Stefan D. Leigh,[§] W. Robert Kelly,[‡] and Andrew L. Rukhin[§]

[†]Measurement Services Division, Technology Services and [‡]Analytical Chemistry Division, Chemistry Science and Technology Laboratory and [§]Statistical Engineering Division, Information Technology Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, Maryland 20899-2300, USA

Received July 21, 2009. Revised Manuscript Received September 30, 2009

Two coal certified reference materials (CRMs) for mercury content were blended, and their results checked, resulting in National Institute of Standards and Technology (NIST) traceable standards at mercury concentrations intermediate to parent certified values. How are CRM blends checked? Interval-repeatability is a new statistic that can be employed as a quality control test to identify incorrectly prepared working standards or nonlinear instrument response over a targeted concentration interval based upon the established repeatability limit (r) of the analytical method used. It is applicable to working standards prepared by CRM dilution, as well as those prepared by CRM blending. Interval-repeatability is unique because it measures the precision between standards of different concentrations. For the measurements of standards to comply with the method precision requirements, the absolute value of observed interval-repeatability must be less than or equal to the analytical method's calculated interval-repeatability limit for a given concentration interval. Interval-repeatability can provide valuable information to the CRM user regarding the internal consistency of standards, measurement precision, the success of standard sample preparation, and instrument linearity. The theory behind this approach is discussed, detailed instructions on its implementation are given, and interval-repeatability quality control test data for mercury in coal standards created by the CRM user are presented.

1. Introduction

The NIST “designer method” was introduced in 2007 to facilitate the mixing of National Institute of Standards and Technology (NIST) certified reference materials, that is, Standard Reference Materials (SRMs), of similar matrices in different proportions, as distinct from calibrants prepared from high purity components.^{1,7} Proper blending enables the certified reference material (CRM) user to create a customized continuum of calibrants and quality control test samples in the matrix of interest with uncertainty estimates that are calculable and traceable to parent CRM certified values. Recognizing that even a working dilution of a CRM, is de facto no longer certified, the resulting blend of two CRMs is likewise not certified, but traceable to the certified values of the parent CRMs. However, unlike dilution with noncertified materials, the NIST blending process provides a diluent with a certified value, and a resulting standard, whose value and uncertainty are traceable. Having an understanding of the technical issues detailed in the designer method is a critical prerequisite to successful blending of CRMs. Recognizing that even a well conceived blend could fail, interval-repeatability (I_r) and its associated interval-repeatability limit ($r_{(i,ii)}$) provide a statistical means to verify CRM blends, as well as those prepared from traditional CRM dilution.

I_r is used when standard (A) is mixed with another known material (B) to prepare a new standard (C_j); it is a function of the interval between measured values (I_m) of paired standards, i , ii , and the interval between their corresponding assigned (i.e., certified or calculated) values (I_a). The difference between corresponding measured and assigned value intervals is I_r . For example, for the interval between A and blend C_j , where $A = i$ and $C_j = ii$:

$$I_{a(A,C_j)} = \text{assigned value of } A - \text{assigned value of } C_j \quad (1)$$

$$I_{m(A,C_j)} = \text{measured value of } A - \text{measured value of } C_j \quad (2)$$

$$I_{r(A,C_j)} = I_{m(A,C_j)} - I_{a(A,C_j)} = \text{interval-repeatability} \quad (3)$$

Regardless of the method used to measure the analyte, comparing I_r to its limit ($r_{(i,ii)}$) provides feedback on the success of the newly prepared standard and instrument linearity. In all cases, at least one CRM must be in the working range of the instrument. The uncertainty associated with the standards, that is, the components of I_a and I_m variance, are discussed in the Supporting Information. Key term definitions are summarized in Table 1.

The interval-repeatability limit ($r_{(i,ii)}$) is based upon the established repeatability limit, r , of the analytical method used over a targeted concentration interval. When a reliable estimate of r does not exist for the repeatability conditions employed (e.g., method, matrix, concentration range) then r is experimentally determined by the test laboratory. ASTM

*To whom to address correspondence: E-mail bruce.macdonald@nist.gov. Fax 301.926.4751. Mail Stop 2300.

(1) Kelly, W. R.; MacDonald, B. S.; Leigh, S. D. *J. ASTM Int.* 2007, 4, 2.

Table 1. Nomenclature

A = low-concentration parent CRM
B = high-concentration parent CRM
C_j = the j th standard blend, a CRM user-prepared standard created by blending CRM A and B . Each identified by a subscripted number (C_1, C_2, \dots)
D = nonparent CRM
i = the first standard on a pair that define an interval, typically the CRM with lowest concentration value in the working range of the instrument
i, ii = a pair of standards that define a concentration interval e.g., A, C_1
X_1 = first measurement of a standard
X_2 = second measurement of the same standard
X_{1i} = first measurement of the first of paired standards that define an interval, i.e., i
X_{2ii} = second measurement of the other standard of the interval pair, i.e., ii
\bar{X} = mean concentration
$[]$ = symbol used for concentration, e.g., $[i]$ is concentration of i
r = repeatability limit, “the value below which the absolute difference between two identical tests results obtained under repeatability conditions may be expected to occur with a probability of approximately 0.95 (95 %),” ASTM D456–08 ¹ (“observed r ” is the observed difference between two test results of a material taken under repeatability conditions)
$r_{[i]}$ = the repeatability limit at the concentration of standard i , provides a conservative estimate of $r_{(i,ii)}$, if $[ii] > [i]$
$I_{a(i,ii)} = \bar{X}_i - \bar{X}_{ii}$ = assigned value interval, i.e., concentration interval between assigned values
$I_{m(i,ii)} = X_{1i} - X_{2ii}$ = measured value interval, i.e., concentration interval between measured values
$I_{r(i,ii)} = I_{m(i,ii)} - I_{a(i,ii)}$ = interval-repeatability, the observed difference between test results of two standards, i and ii , corrected for the difference in their assigned values and taken otherwise under the same repeatability conditions as r
$r_{(i,ii)} = [(r_i^2 + r_{ii}^2)/2]^{1/2}$ is the interval-repeatability limit, the value below which the absolute difference between I_r results for standards i and ii may be expected to occur with a probability of approximately 0.95 (95%)
$ I_{r(i,ii)} \leq r_{(i,ii)}$ is the relationship between observed interval-repeatability and interval-repeatability limit when the four specific conditions given in Section 1.3 are met

International defines $r = 1.96 \times 2^{1/2} s_r$, which is approximated as $r = 2.8 s_r$.^{2–4} Here s_r is the standard deviation of test results obtained under repeatability conditions. The acceptable difference in repeatability conditions between I_r and observed r is that r is based on the difference between two measurements taken at the same concentration from a single material, and I_r is the difference between measurements taken from two materials of the same matrix of different, but known, concentrations i.e., standards i and ii . As will be shown, the CRM user is able to both measure I_r for the concentration interval between their standards and calculate a corresponding I_r limit. To meet method precision requirements, the absolute value of the observed interval-repeatability $I_{r(i,ii)}$ must be less than or equal to the analytical method’s interval-repeatability limit ($r_{(i,ii)}$). Hence, $I_{r(i,ii)}$ is analogous to the observed difference between replicates, as $r_{(i,ii)}$ is to the method’s stated repeatability limit, r .

Precision under repeatability conditions for a given analytical method can be a fixed value, if based on homoscedastic (constant variance) data, or it can vary with concentration if based on heteroscedastic data. The absolute value of the

Table 2. Comparison of Homoscedastic and Heteroscedastic Simulated Data^a

interval i, ii		$r_{(i,ii)}$ given: $r = 4.0$			$r_{(i,ii)}$ given: $r = 0.2\bar{X} + 2$	
i	ii	I_a	I_m	I_r	(homoscedastic)	(heteroscedastic)
A	A	0	3.9	3.9	4.0	4.0
A	C_1	15.0	17.9	2.9	4.0	5.7
A	C_2	24.0	28.9	4.9	4.0	6.8
A	B	30.0	31.9	1.9	4.0	7.6

^a Given: $A = 10$; $B = 40$; $C_1 = (50\% A + 50\% B) = 25$; $C_2 = (20\% A + 80\% B) = 34$.

observed difference between two replicates under repeatability conditions is expected to be less than or equal to the method’s repeatability limit at a stated probability. A fundamental feature of repeatability limits is that they apply directly to the dispersion of the distribution of test results under specified repeatability conditions. Simulated data for parent CRMs A and B , blends C_1 (50:50 blend) and C_2 (20:80 blend) with corresponding I_a , I_m , and I_r calculations based upon both homo- and heteroscedastic data are given in Table 2.

1.1. Homoscedastic Case. Using the simulated data from Table 2 with r having a fixed value of 4, a homoscedastic case is presented in Figure 1. For every concentration and matrix for which r is said to apply when r has a fixed value, test result distributions are assumed to be the same except for the mean values of their concentrations. Each standard has a different mean analyte concentration, yet they have the same r value, and therefore represent identical population distributions in terms of shape and spread. In the lower left corner of Figure 1, the first interval between populations to be evaluated is between A and itself. In this case the populations are identical, the concentration interval between the assigned values is zero, and the difference between replicate measurements (X_1 and X_2) is less than or equal to r . The other three standards (C_1 , C_2 , and B) vary according to their respective mean concentrations and measured values. Because precision is the same regardless of concentrations in homoscedastic cases the interval-repeatability limit is equal to the method’s repeatability limit,

$$r_{(i,ii)} = r \quad (4)$$

and as the concentration changes between repeated measurements,

$$|I_{r(i,ii)}| \leq r_{(i,ii)} \quad (5)$$

For illustration purposes, an example of an outlier is included in Figure 1. In this case I_r for interval A, C_2 (4.9) is greater than $r_{(A, C_2)}$ (4.0). The significance of this outlier is discussed in the Conditions section.

1.2. Heteroscedastic Case. When the data are heteroscedastic, r varies with concentration. In this case, r is most often expressed as a function of \bar{X} (the average of replicate measurements), and r is calculated at this average concentration of the sample from which the measurements were made. In the example given in Figure 2, the value of r for each standard is based on the equation $r = 0.2\bar{X} + 2$. When replicate measurements (X_1 and X_2) are made on different standards (X_{1i} and X_{2ii}), simply using the average of X_{1i} and X_{2ii} is insufficient because it will result in an r value that is neither r_i nor r_{ii} . Rather, both r_i and r_{ii} values are needed to characterize the repeatability limit, $r_{(i,ii)}$. If r is based on heteroscedastic data and repeated measurements are taken

(2) ASTM E 456–08, *Standard Relating to Quality and Statistics*; Annual Book ASTM Standards: West Conshohocken, PA, 2009; Vol 14.02.

(3) ISO 3534–2:2006, *Statistics - Vocabulary and Symbols – Part 2: Applied Statistics*; International Organization for Standardization: Geneva, Switzerland, 2006.

(4) ISO/IEC Guide 99–12:2007, *International Vocabulary of Metrology — Basic and General Concepts and Associated Terms*; International Organization for Standardization: Geneva, Switzerland, 2007. Details are available at www.iso.org.

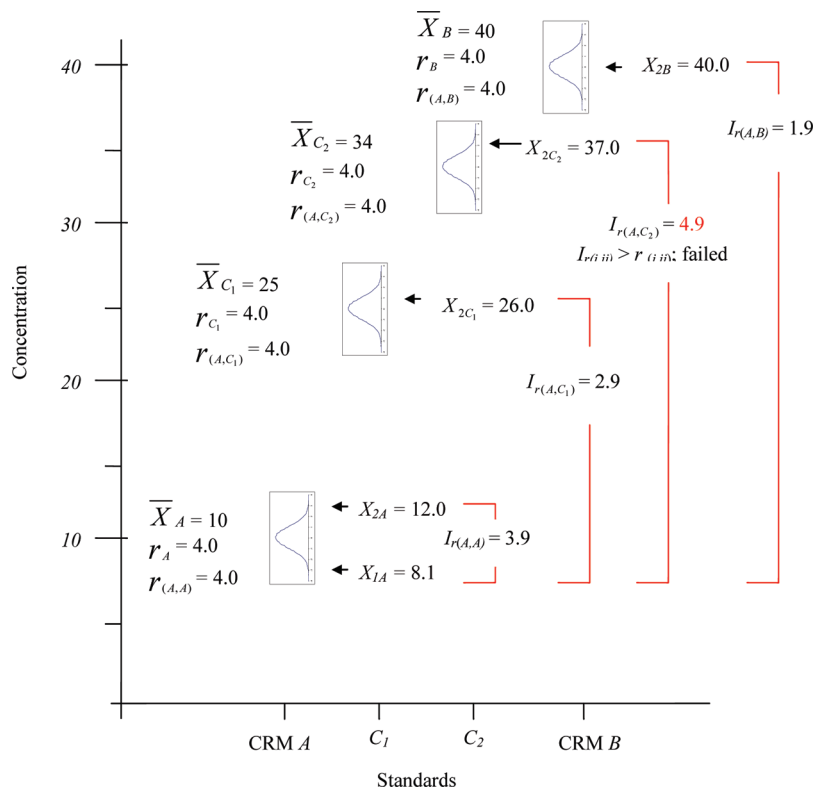


Figure 1. Hypothetical homoscedastic data illustrating $I_{r(i,ii)} \leq r_{(i,ii)}$. Given: $r = 4.0$; $A = 10$; $B = 40$; $C_1 = (50\% A + 50\% B) = 25$; $C_2 = (20\% A + 80\% B) = 34$.

from different standards whose precision is characterized by the same repeatability equation, then $r_{(i,ii)}$ is defined as:

$$r_{(i,ii)} = \sqrt{\frac{r_i^2 + r_{ii}^2}{2}} \quad (6)$$

Then as can be deduced from eqs 5 and 6:

$$|I_{r(i,ii)}| \leq \sqrt{\frac{r_i^2 + r_{ii}^2}{2}} \quad (7)$$

If r_i and r_{ii} are equal, as is the case when the data are homoscedastic, eq 6 simplifies to eq 4. Mathematical proofs of the homoscedastic and heteroscedastic cases are provided in the Supporting Information.

When, instead of a single measurement, each replicate is an average of n measurements, then the right side of eq 7 is multiplied by $n^{-1/2}$ to culminate in the equation:

$$|I_{r(i,ii)}| \leq \frac{1}{\sqrt{n}} \sqrt{\frac{r_i^2 + r_{ii}^2}{2}} \quad (8)$$

Combining multiple measurements into a single result provides a means of increasing sample mass. In the Results section we compare the same experimental data that are calculated using individual ($n = 1$) and pooled measurements ($n = 2$).

1.3. Conditions. Four specific conditions must be met in order for eqs 5, 7, and 8 to be true:

- (1) The assigned values of all nonblended standards are internally consistent.
- (2) Measurements themselves are made in accordance with the method's repeatability limit.
- (3) Prepared standards are sufficiently homogeneous and composed of the intended ratio of components.

- (4) The instrument used is linear in its response over the range of concentrations measured.

If these conditions are met, then the measurements will pass the I_r quality control (QC) test, that is, the absolute value of the measured interval-repeatability (I_r) will be less than or equal to the method's interval-repeatability limit ($r_{(i,ii)}$). Consequently, if I_r is greater than the method's I_r limit ($r_{(i,ii)}$), then the measurements fail the I_r QC test and a problem exists with one or more of the required test conditions. Subsequent investigations of preparation and analysis procedures can be used to identify if the anomaly is most likely due to inconsistency in parent standards, erroneous analysis, errant preparation, a change in linearity, or a "false positive" (at a 95% level of confidence, there is a 1 in 20 chance of a false positive). Recognizing that errors due to condition (1) can be minimized by the type of standards selected for blending (discussed below) and that condition (2) is the minimal requirement for any successful measurement, the I_r QC test is a practical tool to detect condition (3), incorrectly prepared standards, or condition (4), nonlinearity.

In the simulated example of an errant result given in Figure 1, I_r for interval A, C_2 is greater than $r_{(A,C_2)}$ ($4.9 > 4.0$). Linearity and repeatability do not appear to be likely causes because the intervals before and after A, C_2 pass the I_r QC test. Having an internal inconsistency between assigned values of the parent standards is similarly unlikely because the other standards passed the test. Therefore the preparation of blend C_2 needs to be investigated until resolution, or to be discarded.

When CRM blending is done in a manner that minimizes or eliminates concerns about conditions (1), (2), and (3), I_r provides a powerful means for detecting a change in linearity over a targeted concentration range. Knowing an instrument's linear dynamic range is critical to the successful implementation of linear calibration curves and standard additions experiments.

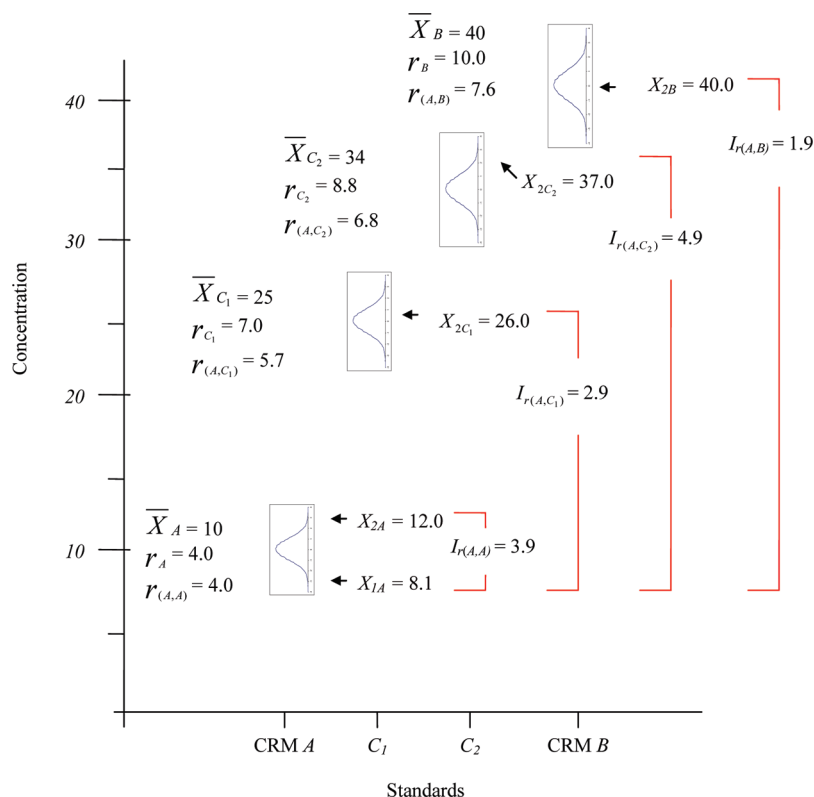


Figure 2. Hypothetical heteroscedastic data illustrating $I_{r(i,ii)} \leq r(i,ii)$. Given: $r = 0.2\bar{x} + 2$; $A = 10$; $B = 40$; $C_1 = (50\% A + 50\% B) = 25$; $C_2 = (20\% A + 80\% B) = 34$.

1.4. Quick and Conservative Estimate of Interval-Repeatability Limit. Even if data are heteroscedastic, the homoscedastic case can provide a quick and conservative estimate of $r_{(i,ii)}$. Referring to eq 4, if one assumes the homoscedastic case, then the analytical method's r at the concentration of the first standard of the interval pair, that is, $[i]$, can be used as a conservative estimate of $r_{(i,ii)}$ for intervals whenever the concentration of the other interval member, $[ii]$ is greater than $[i]$. When the data are assumed to be homoscedastic when in fact they are not, the I_r limit is more conservative, that is, it is more difficult for the observed I_r to be within the limit, and the blend may be fine at an approximate 95% level of confidence even though it fails the test. Although its level of confidence is not as precisely known, the homoscedastic assumption generates a more conservative bar to pass because the repeatability limit at the concentration of A is less than the repeatability limit for the concentration interval A , C_j , that is, $r_{(A,C)}$, and therefore I_r must be smaller yet to pass. Hence r at $[i]$ provides an easy to calculate, initial, pass/fail tool for the CRM-user preparing standards at the bench, regardless if the variance of the data is constant or not. This bench-level screening tool is referred to as an I_r quick initial check (QIC). Blends that appear to fail the I_r QIC need to be evaluated using the I_r QC test (eq 8) to determine if they indeed failed. Data that pass this quick check are certain to pass the I_r QC test.

The I_r QIC was initially referred to in the designer paper: "The accuracy of the blends would need to be checked by measuring the blends and comparing the interval between the measured values of the successive blends to the interval between their corresponding gravimetric value (NIST traceable). If the difference between intervals is less than the repeatability [limit] of the method, then the blends are considered successful."

1.5. Selection of Parent Standards. The accuracy and matrix of parent materials are critical to the success of user prepared standards. Assurance of the absolute accuracy of the parent materials is provided by the level of traceability of the parent CRMs to an appropriate national metrology institute, or other appropriate instantiation of the Systeme International d'Unites (SI). The successful transfer of the CRMs accuracy to the measurement of unknown samples is in large part dependent upon the differences in matrix composition between calibration standards and unknown samples. Potential accuracy pitfalls can be reduced or eliminated when the resulting blends are of the same matrix in terms of the method used for measurement, as the unknowns, and their uncertainties are characterized by the certifying body of the parent CRMs.

2. Experimental Section

The designer method was developed specifically for fuel SRMs, where it is often advantageous to have a CRM at a specific certified value that might not be available. It has been tested with both liquid⁵ and powdered solid fuel SRMs.⁶ The designer method is intended to be used in conjunction with the I_r QC test

(5) Barker, L. R.; Kelly, W. R.; Guthrie, W. F. *Energy Fuels* **2008**, *22*, 2488–2490.

(6) Kelly, W. R.; MacDonald, B. S.; Leigh, S. D.; Eichenbaum, L.; Lawrenz, D.; Marsh, M. *Determination of Sulfur (S) and Mercury (Hg) in Coal: Preparation of NIST Traceable Fossil Fuel Standards with Concentrations Intermediate to SRM Values*. Presented at the ASTM D5 Committee on Coal and Coke Seminar, Norfolk, VA, May 2007.

(7) Kelly, W. R.; MacDonald, B. S.; Leigh, S. D., *A Method for the Preparation of NIST Traceable Fossil Fuel Standards with Concentrations Intermediate to SRM Values (DVD)*, NIST Special Publication SP260–167, September, 2007. This SP260–167(DVD) is free and available through the SRM website www.nist.gov/srm.

Table 3. Mercury Data for Blends of Coal SRMs 2693:2692b

ratio 2693:2692b	nominal % SRM 2693	std	run No.	measured value (mg/kg)	assigned value (mg/kg)	observed r (mg/kg)	method $\pm r$ (mg/kg)	interval i,ii	$n = 1$ interval data (mg/kg)				$n = 2$ interval data (mg/kg)			
									I_m	I_a	I_r	$\pm r_{(i,ii)}$	I_m	I_a	I_r	$\pm r_{(i,ii)}$
1:0	100.0	A	1	0.033	0.0373	-0.003	0.010	A,A	-0.003	0.000	-0.003	0.010	-0.030	-0.016	-0.013	0.007
1:0	100.0	A	2	0.036	0.0373	-0.002	0.011	A,C ₁	-0.030	-0.016	-0.014	0.011	-0.030	-0.016	-0.013	0.007
5:1	83.3	C ₁	1	0.063	0.053	-0.002	0.011	A,C ₁	-0.030	-0.016	-0.014	0.011	-0.030	-0.016	-0.013	0.007
5:1	83.3	C ₁	2	0.065	0.054	-0.002	0.011	A,C ₁	-0.030	-0.016	-0.014	0.011	-0.030	-0.016	-0.013	0.007
4:1	80.0	C ₂	1	0.064	0.057	0.010	0.011	A,C ₂	-0.031	-0.020	-0.011	0.011	-0.024	-0.020	-0.004	0.008
4:1	80.0	C ₂	2	0.054	0.057	0.010	0.011	A,C ₂	-0.021	-0.020	-0.001	0.011	-0.020	-0.020	-0.004	0.008
2:1	66.7	C ₃	1	0.072	0.069	0.003	0.012	A,C ₃	-0.039	-0.032	-0.007	0.011	-0.037	-0.032	-0.005	0.008
2:1	66.7	C ₃	2	0.070	0.069	0.003	0.012	A,C ₃	-0.037	-0.032	-0.005	0.011	-0.037	-0.032	-0.005	0.008
1:1	50.0	C ₄	1	0.077	0.085	-0.005	0.013	A,C ₄	-0.044	-0.048	0.004	0.012	-0.045	-0.048	0.003	0.008
1:1	50.0	C ₄	2	0.082	0.085	-0.002	0.013	A,C ₄	-0.049	-0.048	-0.002	0.012	-0.045	-0.048	0.003	0.008
1:2	33.3	C ₅	1	0.096	0.102	-0.002	0.014	A,C ₅	-0.063	-0.065	0.001	0.012	-0.063	-0.065	0.002	0.008
1:2	33.3	C ₅	2	0.098	0.102	-0.002	0.014	A,C ₅	-0.065	-0.065	-0.001	0.012	-0.063	-0.065	0.002	0.008
1:4	20.0	C ₆	1	0.120	0.114	0.009	0.015	A,C ₆	-0.087	-0.077	-0.010	0.013	-0.081	-0.077	-0.004	0.009
1:4	20.0	C ₆	2	0.111	0.114	0.009	0.015	A,C ₆	-0.078	-0.077	-0.001	0.013	-0.081	-0.077	-0.004	0.009
1:5	16.7	C ₇	1	0.126	0.117	0.014	0.015	A,C ₇	-0.093	-0.080	-0.013	0.013	-0.085	-0.080	-0.005	0.009
1:5	16.7	C ₇	2	0.113	0.117	0.014	0.015	A,C ₇	-0.080	-0.080	0.000	0.013	-0.085	-0.080	-0.005	0.009
0:1	0.0	B	1	0.136	0.1333	0.006	0.016	A,B	-0.103	-0.096	-0.007	0.013	-0.098	-0.096	-0.002	0.009
0:1	0.0	B	2	0.130	0.1333	0.006	0.016	A,B	-0.097	-0.096	-0.001	0.013	-0.098	-0.096	-0.002	0.009

and NIST Special Publication SP260–167, a DVD that includes the designer paper in a noncopyrighted format and instructional videos.

CRM blending has particular time frame restrictions and mass requirements. The traceable values of CRM blends are valid, within their uncertainty, provided CRM blends are used by the time stipulated in the Instructions for Use section of their parent CRMs certificates of analysis. Both solid and liquid CRM blends are to be blended when needed. Blends of powdered reference material are to be prepared on an aliquot-by-aliquot basis only. A sufficient quantity for multiple runs of a liquid standard may be produced in a single blend. No provision is given, or implied, for the long-term storage of blends. The recommended minimum sample mass (MSM) given in the CRM certificate applies to the mass of each replicate measurement of the parent CRMs and subsequent blends. If the MSMs for the parent CRMs are different, then the higher MSM applies to their blends. When the ideal sample mass of the instrument to be used is less than the recommended MSM of either parent CRM, combining multiple (n) measurements each at a mass suitable for the test instrument into a single result provides a practical means of increasing the total sample mass for a replicate. To accurately propagate the uncertainties of the parent CRMs to their blends' assigned values, all results of the blend must be based on a mass equal to or greater than the parents' MSM. However, the mass of each binary component used to create a given blend does not have to be greater than the MSM so long as the resulting blend passes the I_r QC test. While the investigation of I_r for all blend combinations is valuable, the I_r QC test focuses only on the sequence of intervals between the CRM parent with the lowest concentration in the working range of the instrument and each successive blended standard, and culminates with the interval with the other parent (higher concentration) CRM. The basic instructions to the laboratory analyst for blends of liquid and solid (powdered) CRMs are detailed in the Supporting Information.

2.1. I_r QC Test for CRM Blends. After measuring each blend and their parent CRMs, these data will be used to perform an I_r QC test to detect an incorrect blend, or an area of nonlinearity using the analytical method's repeatability limit. It is important that the test instrument have a reasonable calibration, however because this test is based solely on differences the instrument does not require a new recalibration to begin. Once the NIST traceable blends have been tested, their data can be used to create a new working calibration line over the linear range of interest. The following procedure summarizes the I_r QC test when at least one parent CRM is in the working range of the instrument: (1) First identify and record the repeatability limit for the analytical method (i.e., acceptable range between replicates) at each concentration measured. (2) Based on the method's scope, note the working range of the analytical method. (3) Organize samples in order of increasing concentration, that is, $A, C_1, C_2, \dots, C_{j-1}, C_j$, and B (if the high level parent is the only SRM in the working range of the instrument, then organize samples in order of decreasing concentration, that is, $B, C_j, C_{j-1}, \dots, C_2, C_1$, and A). (4) Using the actual masses (g) of each component CRM (i.e., A and B), calculate each blend's assigned value. (5) Determine by difference the assigned value interval (I_a) between the parent CRM with the lowest concentration that is in the working range of the instrument and each successive standard by difference. Include all blends (i.e., C_1 to C_j), and parent SRMs. (6) In the same manner as described in step 5, determine the corresponding measured value intervals (I_m), that is, the intervals between the low concentration parent CRM and each successive measured value. (7) *Successful* blending is indicated whenever the absolute value of the observed interval-repeatability, I_r , (difference between corresponding $I_m(i,ii)$ and $I_a(i,ii)$) is less than or equal to the method's predicted interval-repeatability limit, $r_{(i,ii)}$ (see eq 8). If $|I_r| > |r_{(i,ii)}|$, then the analytical method's repeatability limit at $[i]$, provides an quick and effective conservative estimate of $r_{(i,ii)}$ (see Section 1.3).

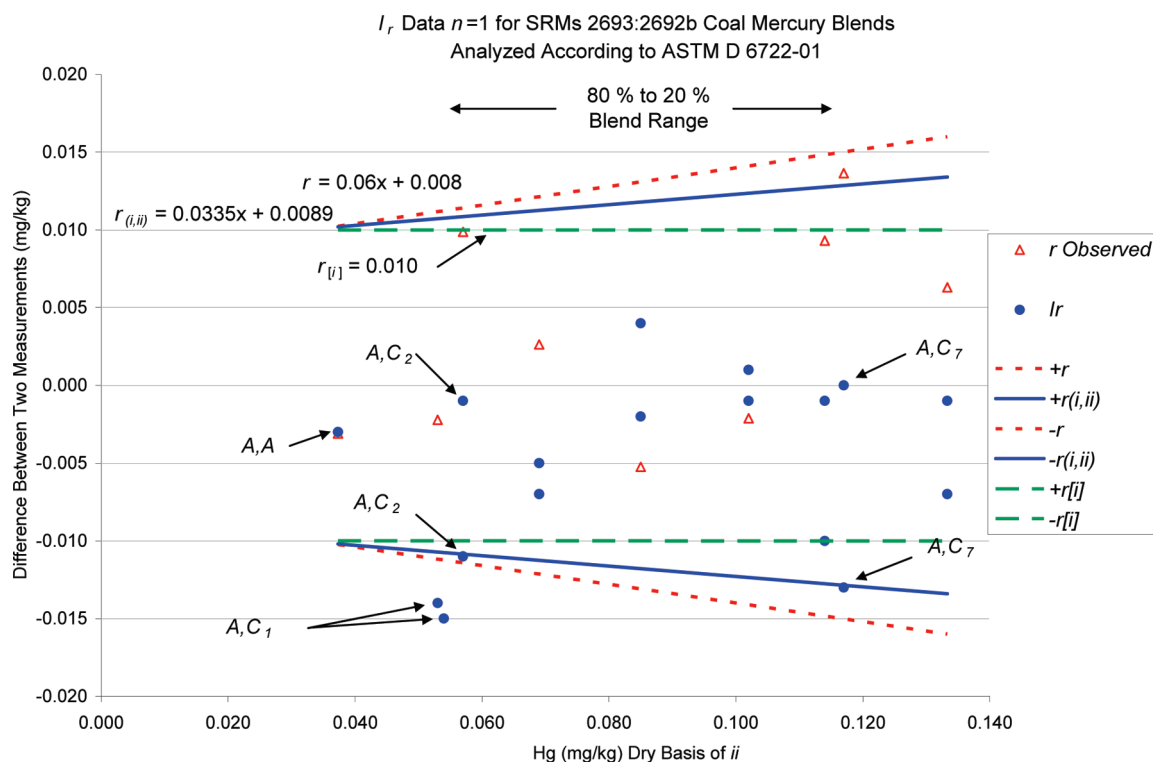


Figure 3. Graph of Table 3 data to illustrate I_r QC Check $n = 1$.

The blends that fail the quick check require evaluation with eq 8 to determine whether they indeed fail the I_r QC test. (8) Unsuccessful blends are to be removed from the sequence of blends, and an updated I_r QC test is performed. Intervals involving parent end members are to be included in the data used in the test, but only the blends in the range of subsequent use are required to pass (this is because, depending on the ratio of the blend and the linearity of the instrument, one or both of the parents, and some of the possible blends, may be out of the linear concentration range of interest). A minimum of two standards in the range of interest, for example, one CRM (A , B , or independent CRM D), and one blend, are needed to perform an I_r QC test. If neither parent CRM (A or B) satisfies this condition, then a nonparent CRM (D) with a certified value in the working range of the instrument is required. The procedure for nonparent CRM D is provided in the Supporting Information. (9) Save a summary of I_r QC test with other details of your blending process for your records. Note: the investigation of I_r with other blend combinations than those mentioned in the I_r QC test can provide additional valuable information.

3. Results

Table 3 contains actual interval-repeatability data on blends of two coal SRMs certified for mercury content. SRM 2693 has a certified mercury content of 0.0373 ± 0.0077 mg/kg (ppm) and SRM 2692b has a certified mercury content of 0.1333 ± 0.0041 mg/kg (ppm).^{8,9} The measurements were determined according to ASTM D 6722–01 Total Mercury in Coal and Coal Combustion Residues by Direct

Combustion Analysis by LECO Technical Services Laboratory using a LECO AMA254 mercury analyzer.^{10,11} The recommended minimal sample mass (MSM) for SRMs 2693 and 2692b is 100 mg. However, to match the instrument's recommended mass-loading more closely, the total mass of coal for each individual mercury measurement was approximately 85 mg. Each parent SRM and 7 blends were analyzed in duplicate for mercury content yielding 18 results. The r for ASTM D 6722–01 varies with concentration (mg/kg) as $r = 0.06\bar{x} + 0.008$. The intervals between the low concentration parent SRM (2693) and each successive standard were calculated according to the I_r QC test, resulting in 17 I_r values ($n = 1$) shown as blue circles in Figure 3. Observed r values for replicate measurements taken at each of the nine concentrations are depicted as red triangles. I_r values in all figures and tables are expressed as positive and negative integers, rather than as absolute values, so that any trends associated with their signs would be visible.

In Figure 3 one can see that all replicate measurements at the same concentration were made in accordance with the method's repeatability limit specification. The first I_r value labeled A,A in Figure 3 illustrates that while I_r and r are identical in theory for two measurements taken at the same concentration (i.e., same assigned value), they can be essentially equivalent in practice too. The overlap of observed r (triangle) and I_r (circle) at A,A is not perfect because during weighing of blend components, there is variation from their target mass that results in slightly different assigned values being calculated.

To perform the I_r QC test, I_r is determined for each of the eight intervals between the low level parent SRM (2693) and

(8) Certificate of Analysis for NIST SRM 2693. Available at https://www-s.nist.gov/srmors/certificates/view_cert2gif.cfm?certificate=2693.

(9) Certificate of Analysis for NIST SRM 2692b. Available at https://www-s.nist.gov/srmors/certificates/view_cert2gif.cfm?certificate=2692b.

(10) ASTM D 6722–01, *Total Mercury in Coal and Coal Combustion Residues by Direct Combustion Analysis*; Annual Book of ASTM Standards; West Conshohocken, PA, 2002; Vol 05.06.

(11) Certain commercial equipment, instruments, or materials are identified in this work to specify adequately the experimental procedure. Such identification does not imply recommendation or endorsement by the NIST nor does it imply that the materials or equipment identified are necessarily the best available for this purpose.

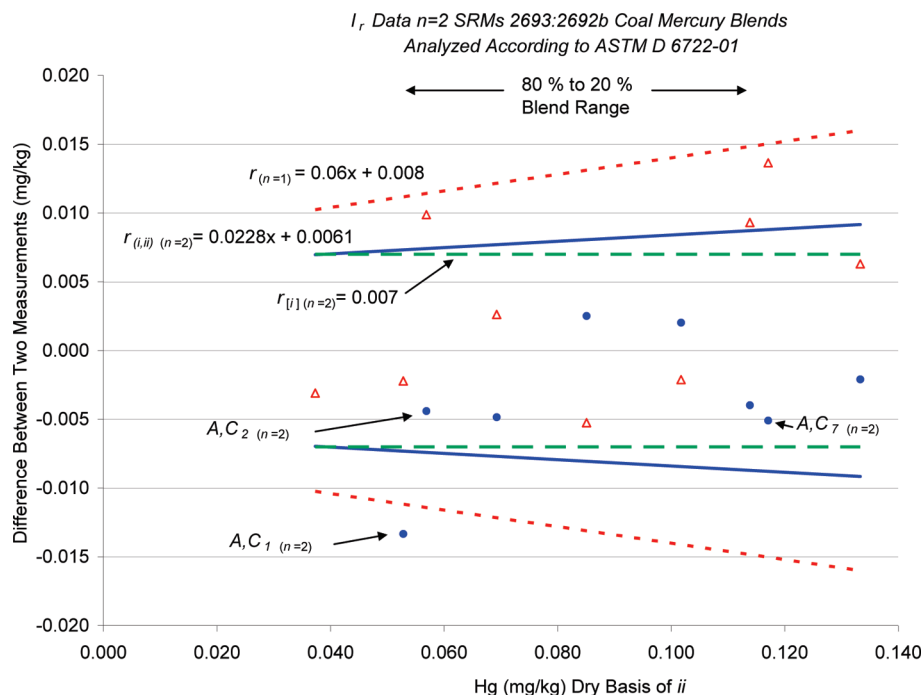


Figure 4. Graph of Table 3 data to illustrate I_r QC check $n = 2$.

each successive standard. Except for interval A, C_1 (between SRM 2693 at 0.0373 mg/kg and the user-prepared 5:1 blends at 0.53 and 0.54 mg/kg), all I_r data passed the I_r QC test (i.e., each I_r is less than or equal to its corresponding $r_{(i,ii)}$ value that is based upon the r for ASTM D 6722–01 and eq 8). Hence, all user-prepared standards in the 4:1 to 1:4 range recommended in the designer method passed the I_r QC test. Therefore, at the stated approximate level of probability of the methods repeatability limit, one can conclude that parent standards are internally consistent, measurements are repeatable, derived standards were prepared successfully, and the instrument is linear in its response over the 4:1 to 1:4 concentration range. However, the mass of the individual blends (85 g) is not sufficient to meet the MSM (100 g) required for traceability to the certified values of SRMs 2693 and SRM 2692b.

To increase the sample mass of replicates for each standard in Table 3, individual runs were averaged so that $n = 2$. This reduced the number of results from 18 to 9 and the corresponding number of intervals from 17 to 8. Treating two runs as one replicate also helps to clarify the data by effectively reducing random error. For example in Figure 3 ($n = 1$), one of the two I_r points for intervals A, C_2 and A, C_7 appears to just barely pass the I_r QC test. With $n = 2$, the average of both I_r points for intervals A, C_2 and A, C_7 are within their correspondingly tighter interval-repeatability limits in Figure 4. In addition to the $r_{(i,ii)}$ limit being tighter and the I_r data being less variable, MSM requirements are met when n is increased.

4. Conclusion

Interval–repeatability (I_r) is a statistic that can be employed by users of CRMs to verify the successful preparation of working standards. Once one can assume that the standards are successfully prepared, then I_r also provides a sleuthing tool capable of identifying areas of nonlinearity. However, the I_r QC test is only applicable for checking CRM blends when used in conjunction with a carefully considered and implemented blending process. Issues associated with selection of

parent CRMs, calculation of uncertainty, miscibility of liquid CRMs, the type of analytical methods that are amenable to blends of liquid and solid CRMs, and other related topics are considered in NIST Special Publication SP260–167.⁷ In addition, a downloadable NIST spreadsheet is planned. The Supporting Information for this paper includes a list of fossil fuel SRMs NIST has successfully tested for creating traceable standards at a continuum of concentrations intermediate to SRM mercury and sulfur certified values.

The comparison of the observed interval-repeatability to the method's calculated interval-repeatability limit is based upon the established repeatability limit (r) of the analytical method used, a criterion familiar to laboratory scientists and technicians worldwide. While a statistic based on r is not the only possible approach, I_r provides a uniform, conceptually familiar means of identifying incorrect blends or instrument nonlinearity.

Acknowledgment. The authors would like to acknowledge Liliane Eichenbaum, of the LECO Technical Services Laboratory for her blending and analysis of coal SRMs for mercury content; Dennis Lawrenz, Director of the LECO Technical Services Laboratory, and Mason Marsh, LECO Product Specialist for Organic Analysis for their kind assistance with LECO Corporation's technical collaboration; William F. Guthrie, of the NIST Statistical Engineering Division, for his technical consultation; and David L. Duewer, of the NIST Analytical Chemistry Division for his detailed review.

Supporting Information Available: The following is a list of supporting information related to this paper: (1) interval-repeatability: I_a and I_m components of uncertainty, (2) interval-repeatability: homoscedastic case, (3) interval-repeatability: heteroscedastic case, (4) guide for the preparation of liquid CRM blends, (5) guide for the preparation of solid (powdered) CRM blends–aliquots only, (6) I_r QC test for blends requiring a nonparent CRM D , and (7) fossil fuel SRMs successfully tested for blending within their matrix for sulfur and mercury. This material is available free of charge via the Internet at <http://pubs.acs.org>.